
RPM 2017
Trees in the R environment
DIY analysis with open source software

Modelers are builders and use tools

Per Merriam-Webster a ***tool*** is

- A handheld device that aids in accomplishing a task




- An apparatus used in the practice of a vocation or profession

The *R Environment* is a collection of tools

CRAN The Comprehensive R Archive Network

R-bloggers

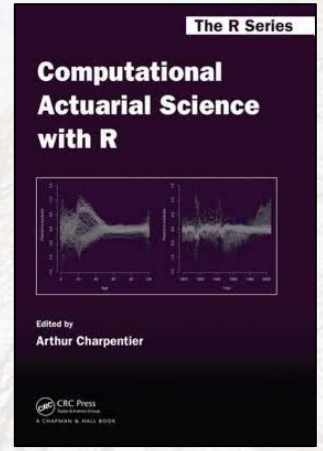
Microsoft Revolution Analytics

Mailing Lists

Please read the instructions below

[R-sig-ins]

R in Insurance 2016



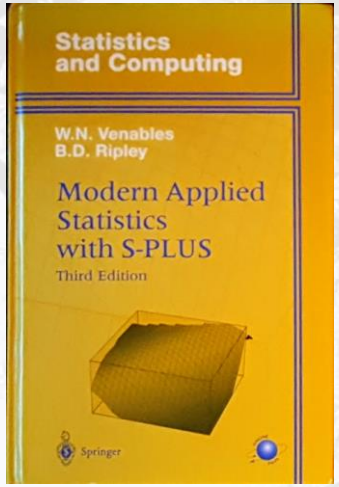
Shiny

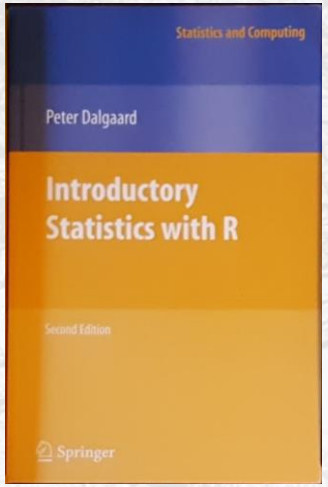
R Markdown

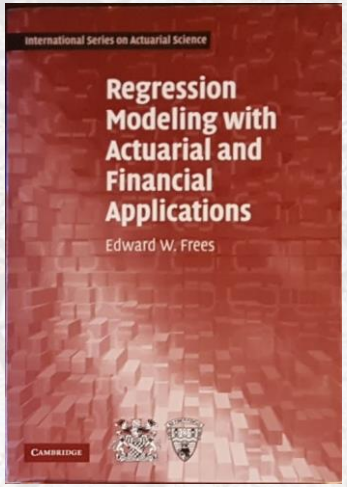
GitHub, Inc. [US] https://github.com

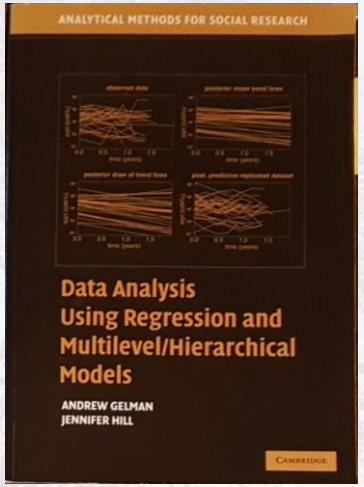
London R We're 2,110 R Users

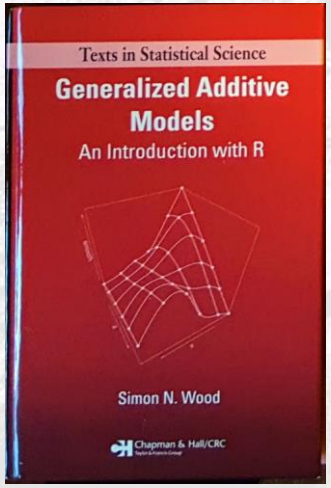
Learn statistics with R

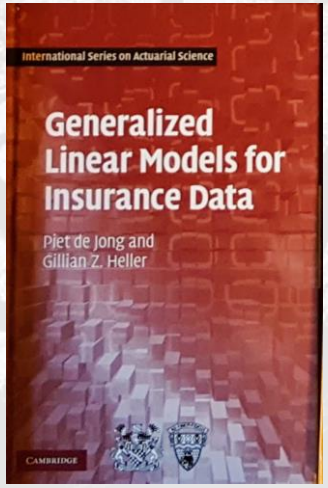


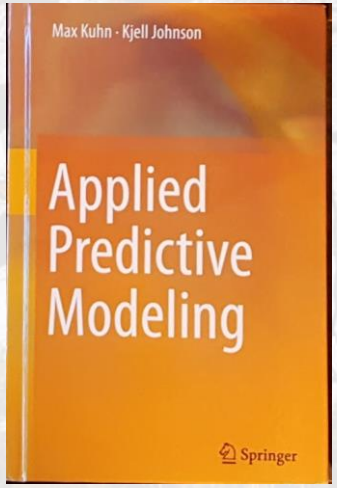


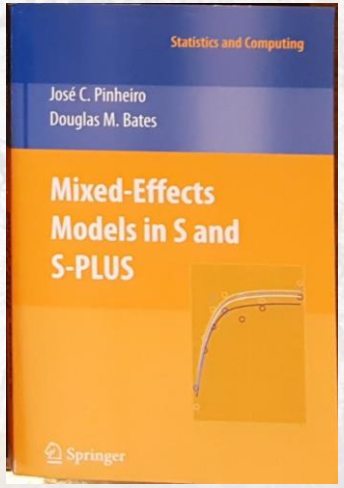












rpart help

```
> ?rpart
```

R: Recursive Partitioning and Regression Trees ▾ Find in Topic

rpart {rpart} R Documentation

Recursive Partitioning and Regression Trees

Description

Fit a `rpart` model

Usage

```
rpart(formula, data, weights, subset, na.action = na.rpart, method,
      model = FALSE, x = FALSE, y = TRUE, parms, control, cost, ...)
```

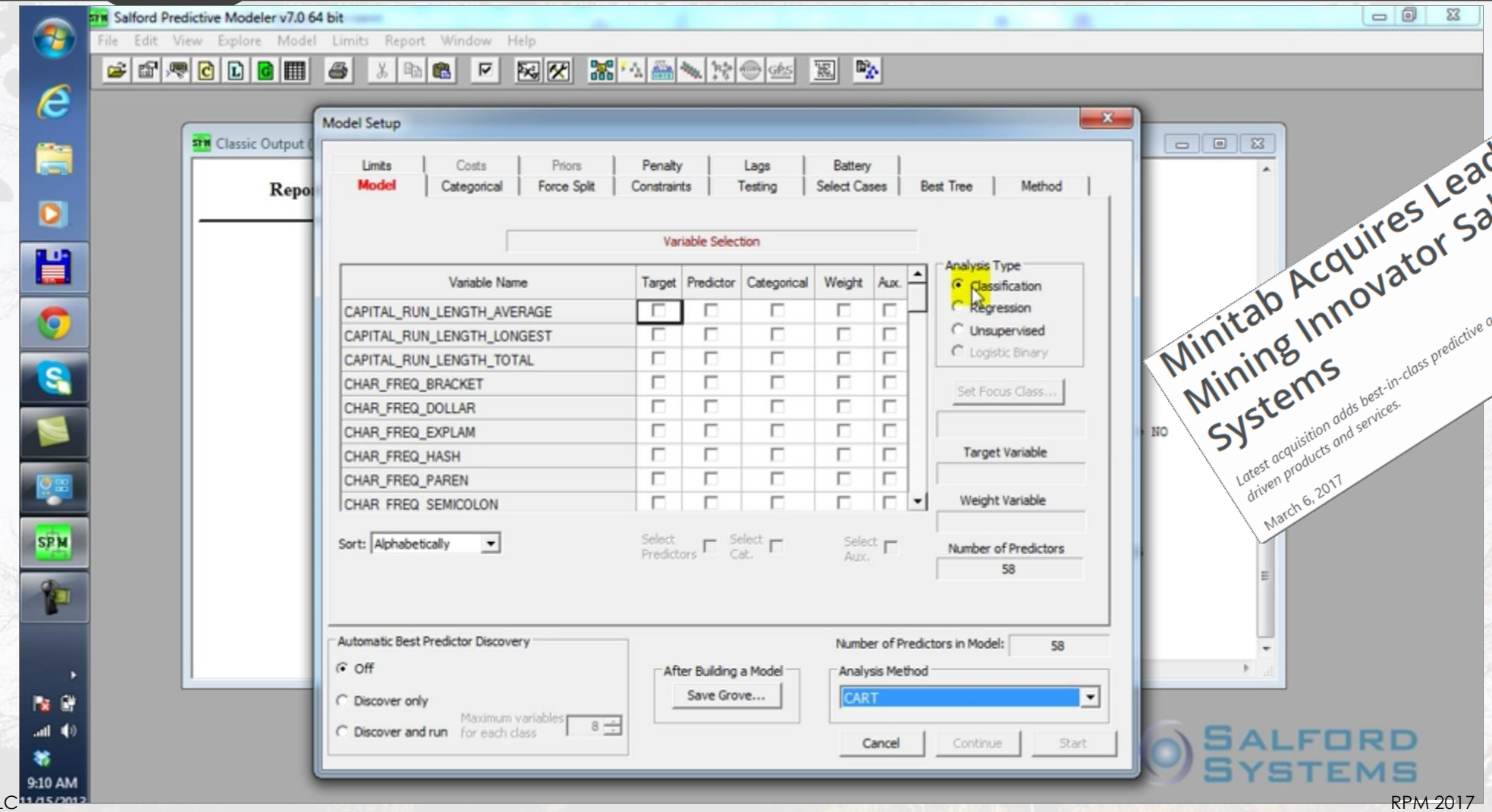
Arguments

`formula`	a formula, with a response but no interaction terms. If this a a data frome, that is taken as the model frame (see `model.frame`) .
`data`	an optional data frame in which to interpret the variables named in the formula.
`weights`	optional case weights.
`subset`	optional expression saying that only a subset of the rows of the data should be used in the fit.
`na.action`	the default action deletes all observations for which `y` is missing, but keeps those in which one or more predictors are missing.
`method`	one of `"anova"`, `"poisson"`, `"class"` or `"exp"`. If `method` is missing then the routine tries to make an intelligent guess. If `y` is a survival object, then `method = "exp"` is assumed, if `y` has 2 columns then `method = "poisson"` is assumed, if `y` is a factor then `method = "class"` is assumed, otherwise `method = "anova"` is assumed. It is wisest to specify the method directly, especially as more criteria may added to the function in future.
	Alternatively `method` can be a list of functions named `init`, `split` and `eval`. Examples are given in the file 'tests/usersplits.R'

Recursive PARTitioning

Because CART is the trademarked name of a particular software implementation of these ideas, and tree has been used for the SPlus routines of Clark and Pregibon [2] a different acronym — Recursive PARTitioning or rpart — was chosen. It is somewhat humorous that this label "rpart" has now become more common than the original and more descriptive "cart", a testament to the influence of freely available software.

An alternative



Minitab Acquires Leading Data Mining Innovator Salford Systems

Latest acquisition adds best-in-class predictive analytics tools to Minitab's array of data-driven products and services.

March 6, 2017

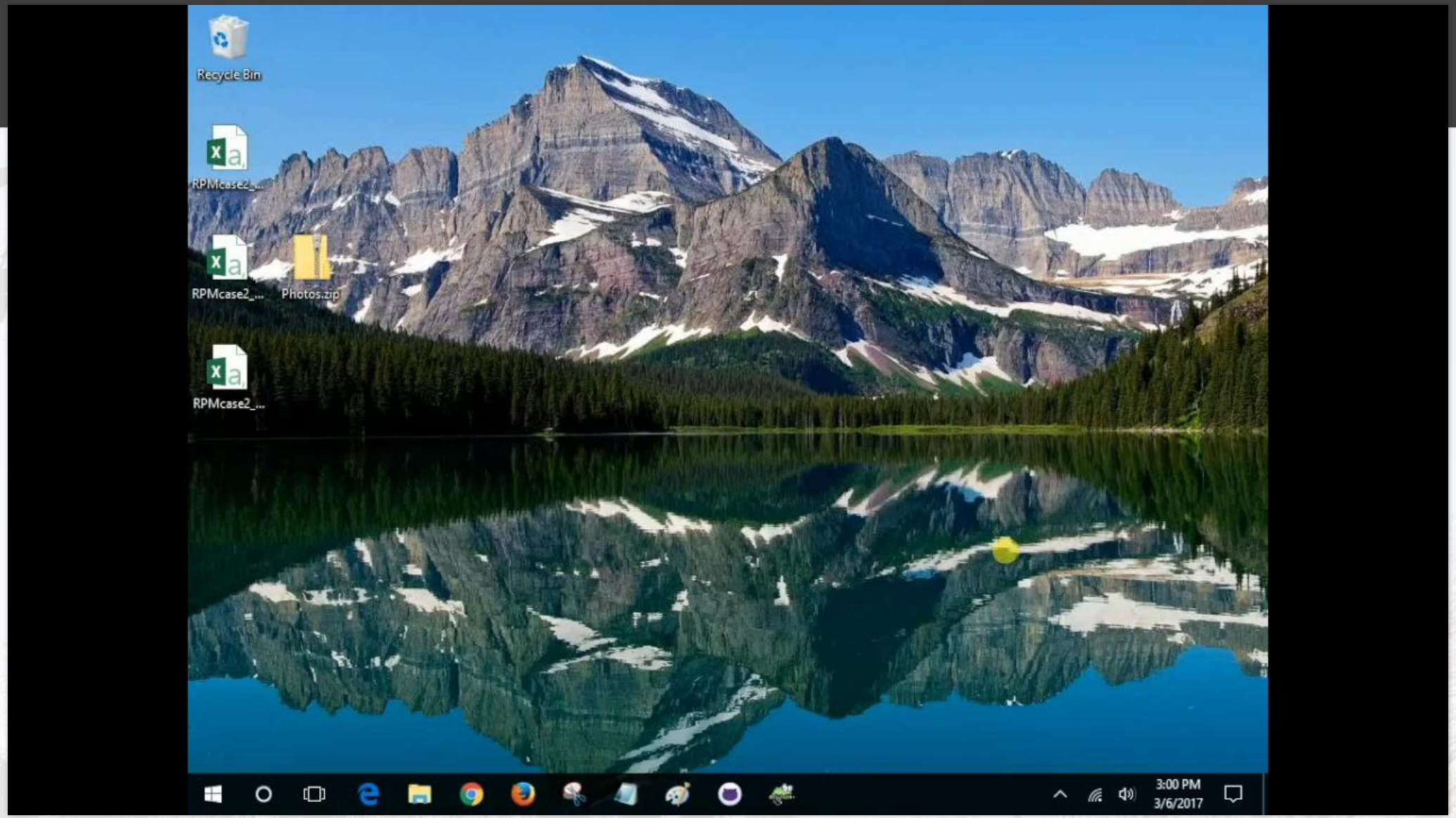
rpart in action/detail

- Start RStudio
- Load the rpart library
- Load the data
- Do your thing

Run R/Rstudio/Rattle from USB drive

- Copy .Rprofile to folder containing R code
  - Easy-Tree-sy\Rattle\code\Rattle.R
- Double-click the R file
  - Starts RStudio
  - Opens file to edit
  - Run code with RStudio's Run button
- Install the Rattle package using RStudio's gui
  - "Depends" is necessary
  - "Suggests" is nice but can be overkill
- Load the library
- Start Rattle
- Keep installing necessary packages until Rattle runs without error

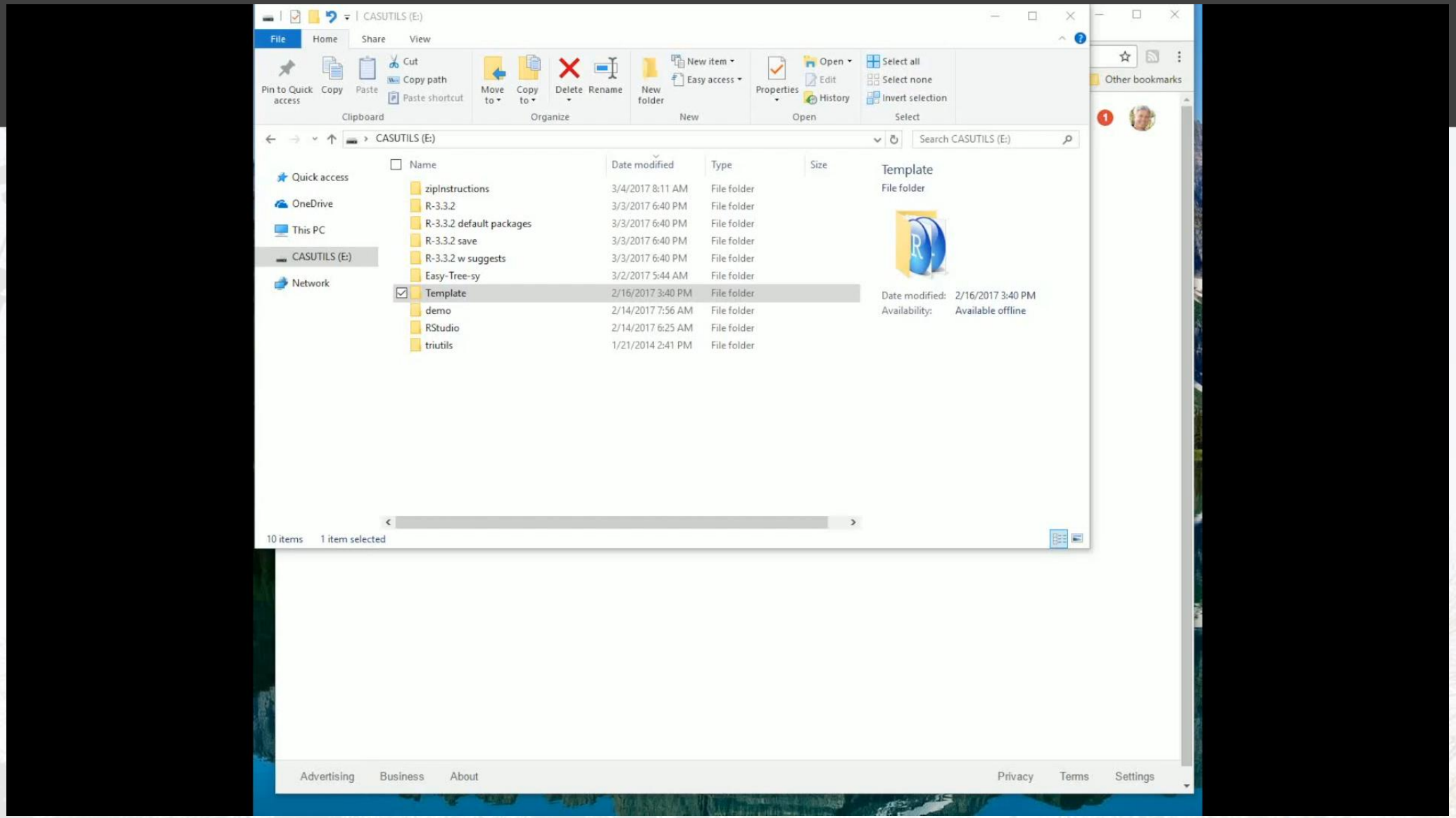
Run R/Rstudio/Rattle from USB drive



Split the data with a *shiny* app

- Split policyholder universe into three tranches
  - Test
  - Train
  - Validate
- Premise: for each individual policyholder, want all policies in same tranche
- Solution: Use the leading digit of policy number to determine which policies go into which tranche
  - E.g., All policy numbers P1xxxxxx go into "Test"
- Online app at https://trinostics.shinyapps.io/SplitPolicies/

Split the data with a *shiny* app



Summary

- Cutting edge apps these days tend to be made from open source tools
- For hands-on learners and modelers, the R environment is growing, maleable
- The insurance industry is under-served at this point in time

Contact me if questions

Dan Murphy
dmurphy@trinostics.com
Trinostics LLC, San Francisco

https://www.triknowbits.com/

https://cran.r-project.org/

https://www.rstudio.com/

https://www.salford-systems.com/